

PREPARED FOR SUBMISSION TO JHEP

General Relativity Tutorials 2026

Lucas T. Y. Leung^a

^a*Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK*

E-mail: lucas.leung@physics.ox.ac.uk

ABSTRACT: A set of notes for tutorials for General Relativity B5 in Oxford.

Contents

0	Introduction	1
1	From special to general relativity	2
2	Manifolds, vectors and tensors	5
2.1	Manifolds	5
2.2	Vectors	7
2.3	Covectors and tensors	11
2.4	Tensor fields	14
3	Metrics	16
3.1	A first look at metrics	16
3.2	Signature of the metric	17
3.3	Embedding metrics	18
4	Connections and Geodesics	18
4.1	Connections I — as a way to fix differences of vectors between two points	19
4.2	Connections II — as transformations of tetrads	20
4.3	Intrinsic derivatives and geodesics	23
4.4	Variational methods and geodesics	25
4.5	Covariant derivatives and Koszul connections	29
5	Curvature	29
A	Maps between manifolds	29

0 Introduction

Welcome to general relativity!

This is not an easy subject. Many shy away from this course as this is perhaps one of the most mathematically challenging topics in a physics degree. But I promise it will be worth it. These set of notes therefore serve as a nice basis for our discussion — they are often riddled with mistakes so I would expect you to spot them!

1 From special to general relativity

The centre point of special relativity is that physics in any inertial frames are the same. In particular, we can recall the two postulates in special relativity.

Postulate 1.1 (The first postulate of relativity). The laws of nature are the same in all inertial reference frames.

Of course, we also know that there is fundamental speed central to physics and this speed — namely the speed of light — is constant to all observers. This motivates the second postulate of relativity.

Postulate 1.2 (The second postulate of relativity). The speed of light in vacuum, c , is the same in all inertial reference frames.

The key point is we needed something known as an **inertial reference frame** — this is a **system of reference** which is not acted upon by external forces and proceeds with constant velocity. The theory of special relativity is developed with this principle in mind, and we have studied this in detail in B2 with Caroline.

The next obvious question is what happens to non-inertial frames. To this let us actually think about the effects of *gravitational fields*. Gravitational fields have the property that no matter the mass of the test object, all bodies will move in them in the same manner. This is encoded in the Weak Equivalence Principle, stated normally as follows.

Postulate 1.3 (Weak Equivalence Principle). The trajectory of a freely falling test body depends only on its initial position and velocity, and is independent of its composition.

This creates a big problem — if we wish to measure the effects of a gravitational field, we must use some test object and something ‘stationary’ to compare the effects, but there are simply no such ‘stationary’ objects as all objects are affected by the gravitational field in the same manner! These stationary objects can only be forced to exist by introducing non-inertial frames (say for example, imagine you are superman and holding a ball in place next to a black hole). So if we work locally, the properties of the motion in a non-inertial system are the same as those in an inertial system in the presence of a gravitational field. Einstein extended the equivalence principle stated above to the now famous *principle of equivalence*, for which we can state as follows.

Postulate 1.4 (Einstein’s Equivalence Principle). The following statements hold true.

- (i) The WEP is valid.
- (ii) In a local inertial frame, the results of all gravitational experiments will be indistinguishable from the results of the same experiments performed in an inertial frame in Minkowski spacetime.
- (iii) The properties of motion in a non-inertial system are **locally** the same as those in an inertial system in the presence of a gravitational field. In particular, a non-inertial reference frame is equivalent to a certain gravitational field.

A constant gravitational field therefore corresponds to a uniformly accelerated reference frame; and a variable field corresponds to a non-uniformly accelerated reference frame. As theoretical physicists, the clear next step is to find the correct mathematical tool to describe this. Let us recall what happens in the Minkowski case. In an inertial reference system the spacetime interval is given by the relation,

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 . \quad (1.1)$$

Let us think about what happens when we transform to a non-inertial frame. Suppose we have a non-inertial system describing a rotating frame at uniform angular velocity Ω . Clearly we can take the transformation,

$$t = t' , \quad \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos \Omega t & -\sin \Omega t \\ \sin \Omega t & \cos \Omega t \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} , \quad z = z' , \quad (1.2)$$

which now gives the following spacetime interval (which we will, with anticipation, call the **metric**),

$$ds^2 = -(c^2 - \Omega^2(x'^2 + y'^2))dt^2 + dx'^2 + dy'^2 + dz'^2 - 2\Omega y' dx' dt + 2\Omega x' dy' dt . \quad (1.3)$$

We can no longer write this as a sum of squares of coordinate differentials. This means the metric appears as a quadratic form of general type in coordinate differentials, i.e.

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu . \quad (1.4)$$

So perhaps we should do the same for gravitational fields too. This means the correct mathematical object to look at is $g_{\mu\nu}$ — the metric of the spacetime. But we aren't done. Notice the word **locally** in EEP. The equivalence principle only states that in a sufficiently small region of spacetime the gravitational effects are the same as a non-inertial frame. There is, however, a fundamental difference between gravitational effects and non-inertial systems. At infinite distances away from the source of the gravitational effects, our physical intuition tells us that the gravitational field goes to zero. After all, *locality* remains in play, and things that are far far away should not affect the local dynamics. This however, does not affect non-inertial frames — at infinity ¹, the effects will be the same.

How does this translate to our mathematical structure, namely, the metric? Since a non-inertial frame has globally the same structure (everywhere), it must be true that we can make a global change of coordinates to get back to an inertial reference frame. This is indeed the case in the uniformly rotating frame we have discussed above — we have made a global change of coordinates to obtain the new metric. In mathematical terms, non-inertial metrics, $h_{\mu\nu}$, are related to the Minkowski metric by a change of coordinates $x^\mu \mapsto y^\mu$,

$$ds^2 = h_{\mu\nu} dy^\mu dy^\nu = g_{\mu\nu} dx^\mu dx^\nu . \quad (1.5)$$

¹I am giving a very loose physical definition of infinity here. Think of it as spatial infinity, but we should probably do a proper characterisation of infinities later.

This is however not true for gravitational metrics — we cannot make a global coordinate transformation as the gravitational field itself goes to zero at infinity! This suggests that the metric associated to gravitational fields cannot be transformed away to a Minkowski one by a simple coordinate transformation.

This forms the basis of general relativity. We want to understand how gravitational effects come into play, and it turns out that they affect the spacetime interval, or metric, in a very non-trivial way. This is what people mean when they say ‘gravity curves spacetime’. But now we have a better understanding, and as theorist we shall now proceed to understand the mathematical structures that describes this phenomenon as well as the physical consequences of this mathematical description.

So we have a wish-list. A wish-list for all the mathematical structures we want for describing gravity in spacetime. Clearly, the spacetime interval ds^2 must come into play, so let me list the ingredients down below.

1. **A space (manifold M).** Clearly, before we even begin, we will need some description of space that describes our world and locally admits some inertial frame (coordinates) so we can label events. Mathematically this is known as a differentiable manifold, and this will be the subject of the first tutorial.
2. **A measure on the space for spacetime intervals (metric g).** As mentioned we will need the metric — this object must be able to intrinsically measure gravitational effects, and we will see how notions of connections and curvatures will come into play and tell us something about gravity affects spacetime.
3. **Descriptions of fields (tensor fields).** We will of course also need some notion of vectors and tensor to describe physical fields in this space. This turns out to be in general tensor fields, and we will need to set up properly how they work so we have consistent mathematical descriptions of fields in spacetime.

The fundamental mathematical object in general relativity turns out to be (M, g) , a manifold equipped with a (Lorentzian metric). In fact, this is the only mathematical object needed to describe spacetime ². This is the second important principle in GR, known as general covariance.

Postulate 1.5 (General Covariance). The only mathematical structure to describe the physical spacetime is (M, g) .

The special version of this will be that the only mathematical structure needed to describe inertial frames is Minkowski spacetime, with the metric $ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu$.

So here we are. The rest of our discussion should clearly focus on understanding the mathematical structures, and then discuss the physical consequences of these postulates and set-up. General relativity is actually not too hard — once you wrap your head around

²Here I mean the physical one, which inconveniently sometimes people call (M, g) , the mathematical object, spacetime too.

the mathematics, everything will be straight-forward, but the mathematics, or differential geometry, is of course quite difficult! In fact, the whole field of differential geometry, one might argue, has been sped up by the development of the theory of relativity. But don't fear — I will be here to join you along the way.

2 Manifolds, vectors and tensors

In this section we begin our study of differential geometry.

2.1 Manifolds

Our discussion in differential geometry begins with manifolds. A manifold is a space which locally looks like an \mathbb{R}^n .

Definition 2.1. A n -dimensional smooth manifold M is a topological space such that there exist pairs (U_α, ϕ_α) called **charts**, where,

- (i) $\bigcup_\alpha U_\alpha = M$, so (U_α) forms an open cover over M .
- (ii) For each α , there is a homeomorphism $\phi : U_\alpha \rightarrow V_\alpha \subset \mathbb{R}^n$, where V_α are open sets in \mathbb{R}^n .
- (iii) For $U_\alpha \cap U_\beta \neq \emptyset$, then $\phi_\beta \circ \phi_\alpha^{-1} : V_\alpha \rightarrow V_\beta$ is called the **transition map** which maps the overlap region to the other chart. The transition map is differentiable (smooth).

We call ϕ_α **charts**, **coordinate systems** or **trivialisations**. The set $\{\phi_\alpha\}$ is called an **atlas** (see Figure 2.1).

So a chart is like a local map — it allows you to give local coordinates at a point.

$$\phi_\alpha(p) = (x_\alpha^1(p), x_\alpha^2(p), \dots, x_\alpha^n(p)). \quad (2.1)$$

Imagine you are walking along Oxford Street in London and you walk by the maps that show the tourist information in the area. You can walk a few blocks along the street and still see a similar map but shifted up a few blocks — but the streets and stations (say, Oxford Circus tube station) that are shown on the two maps must be the same and presented in a similar way (since, clearly, they represent the same buildings)! This means the ‘map’ going between the two charts must be smooth.

Topological spaces and manifolds

We have gone through a bit fast here. Technically, we should start by defining what is a topological manifold. Let me suppose we don't know any topology.

Definition 2.2. A **topological space** $T = (X, \mathcal{T})$ consists of a non-empty set X together with a fixed family \mathcal{T} of subsets of X satisfying,

(T1) $X, \emptyset \in \mathcal{T}$.

(T2) For $A, B \in \mathcal{T}$, $A \cap B \in \mathcal{T}$.

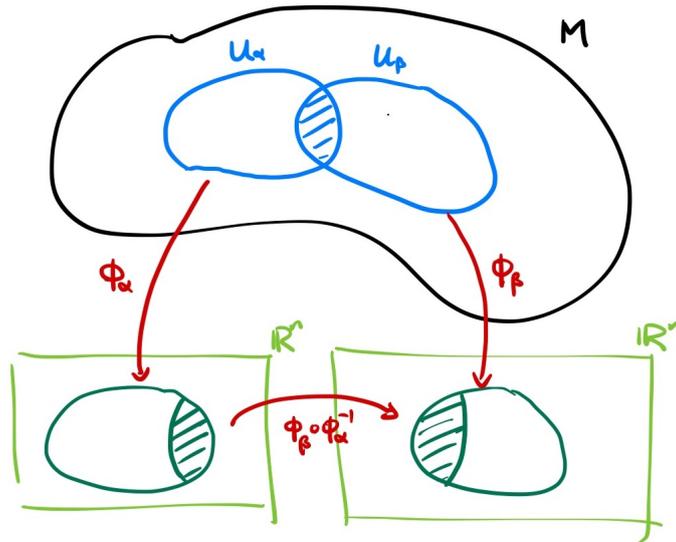


Figure 2.1: A depiction of a smooth manifold.

(T3) For $A_i \in \mathcal{T}$, where $i \in I$ and I can be infinite (I is an indexing set), we have $\bigcup_i A_i \in \mathcal{T}$.

The family \mathcal{T} is called a topology of X and the members of \mathcal{T} are called **open sets** of \mathcal{T} .

Let us also define what a homeomorphism is.

Definition 2.3. A map of topological spaces $f : (X, \mathcal{T}_X) \rightarrow (Y, \mathcal{T}_Y)$ is **continuous** if $U \in \mathcal{T}_Y \Rightarrow f^{-1}(U) \in \mathcal{T}_X$. A **homeomorphism** between topological spaces X and Y is a bijective map $f : X \rightarrow Y$ such that both f and its inverse function f^{-1} exist and are both continuous.

With this we can now define a topological manifold.

Definition 2.4. A n -dimensional topological manifold M is a topological space that satisfies the following properties.

1. M is a **Hausdorff space**: for every pair of distinct points $p, q \in M$, there are disjoint open subsets $U_p, U_q \subset M$ such that $p \in U_p$ and $q \in U_q$.
2. M is **second-countable**: there exists a countable basis for the topology of M .
3. M is **locally Euclidean of dimension n** : each point of M has a neighbourhood, $p \in U_\alpha$ that is homeomorphic to an open subset $V_\alpha \subset \mathbb{R}^n$, i.e. $\phi_\alpha : U_\alpha \rightarrow V_\alpha$ is a homeomorphism.

With a topological manifold, one can then introduce transition maps as before. We can then define smooth structures.

Definition 2.5. Two charts (U_α, ϕ_α) and (U_β, ϕ_β) are said to be **smoothly compatible** if either $U_\alpha \cap U_\beta = \emptyset$ or the transition map $\phi_\beta \circ \phi_\alpha^{-1}$ is a diffeomorphism.

An atlas \mathcal{A} for M is a **smooth atlas** if any two charts are smoothly compatible with each other. We define a **smooth structure** as an equivalence class of smooth atlases which is **maximal**, i.e. it is not properly contained in any larger smooth atlas. A **smooth manifold** is a pair (M, \mathcal{A}) where M is a topological manifold and \mathcal{A} is a smooth structure on M .

A bit of a roundabout way to define everything properly, but here it is.

One could perhaps ask why we need to map open sets to open sets. You should really think of open sets as the basic units in topology — they are the members of topology. To properly define homeomorphisms, we will need to find open sets in both M and \mathbb{R}^n , and we know that the open sets (the basic units) in \mathbb{R}^n are the ‘open sets’ in the normal sense — the interval sets that are unbounded.

Let us have a look at a few examples.

Example 2.1 (\mathbb{R}^n as a smooth manifold). The simplest example is \mathbb{R}^n — the open sets are just the open sets in \mathbb{R}^n and the charts are just the identity maps.

Example 2.2 (S^1 , the circle). The simplest non-trivial example is the circle S^1 . We can define two sets which covers S^1 — $U_N = S^1 \setminus \{N\}$ and $U_S = S^1 \setminus \{S\}$ where N and S are the north pole and south pole respectively (Figure 2.2). We can construct the charts by using the angular coordinate θ : for U_N we map to $\theta_N \in (0, 2\pi)$ and for U_S we have $\theta_S \in (-\pi, \pi)$. We define the transition function as follows:

- For the left side of the overlap the coordinates match, so $\theta_S = \phi_S \circ \phi_N^{-1}(\theta_N) = \theta_N$.
- For the right side of the overlap the coordinates are off by 2π , so $\theta_S = \phi_S \circ \phi_N^{-1}(\theta_N) = \theta_N - 2\pi$.

2.2 Vectors

The next point is to try and understand objects on the manifold as potential candidates to describe our various fields. To do this we will need some extra notions.

To properly define vectors, we will first need a notion of smooth functions on a smooth manifold. The basic logic in dealing with manifolds is this — we intrinsically only have notions of differentiable when we talk about maps from \mathbb{R}^n to \mathbb{R}^m . Therefore, we must define properties of functions with respect to \mathbb{R}^n , and manifolds naturally provide us with charts that maps a local region of M to an open set in \mathbb{R}^n .

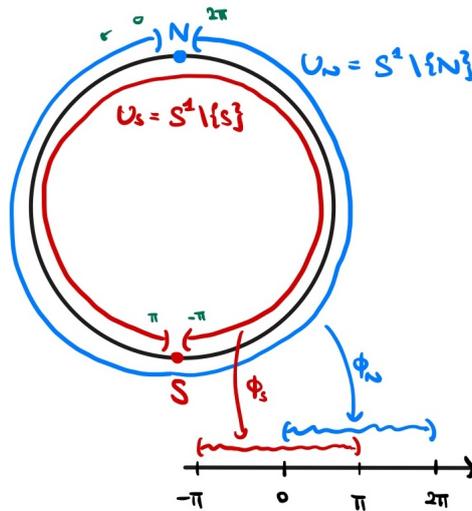


Figure 2.2: The manifold S^1 . The blue and red lines illustrate the two open sets U_N and U_S respectively.

Definition 2.6. A function $f : M \rightarrow \mathbb{R}$ is **smooth** iff for any chart ϕ , $f \circ \phi^{-1} : V \rightarrow \mathbb{R}$ is a smooth function for $V \subset \mathbb{R}^n$.

Note that this gives us a notion of a scalar field!

To define vectors, it is clear that we can consider some notion of a tangent space at a point. But clearly tangent spaces at different points may not align (think of trying to put a flat surface on various points on the surface of a sphere — these flat surfaces do not align). To do this, we will need to first define what curves are on manifolds.

Definition 2.7. A **smooth curve** in a smooth manifold M is a smooth function $\gamma : I \rightarrow M$ where I is an open interval in \mathbb{R} , i.e. $\phi_\alpha \circ \gamma : I \rightarrow \mathbb{R}^n$ is a smooth map for all charts ϕ_α .

Why do we need this notion? Let us think about how we would get tangent vectors in the first place. Suppose we have a function $f : \mathbb{R} \rightarrow \mathbb{R}$. It is clear that a notion of a vector is naturally given by the rate of change of f along \mathbb{R} — this is just the local derivative at $p \in \mathbb{R}$ of the domain. In our current case, we fortunately can construct a function $f \circ \gamma : I \rightarrow \mathbb{R}$ which has some notion of a derivative. This means we can extend the notion of tangent vectors to a curve — we can look at the rate of change of f along a curves in \mathbb{R}^n as the directional derivative $X_p \cdot (\nabla f)_p$ (see Figure 2.3).

Definition 2.8. Let $\gamma : I \rightarrow M$ be a smooth curve without self-intersections and $\lambda(0) = p$ without loss of generality. The **tangent vector to γ at p** is the linear map X_p from the space of smooth functions on M to \mathbb{R} given by

$$X_p(f) = \left\{ \frac{d}{dt} f(\gamma(t)) \right\}_{t=t_0}. \quad (2.2)$$

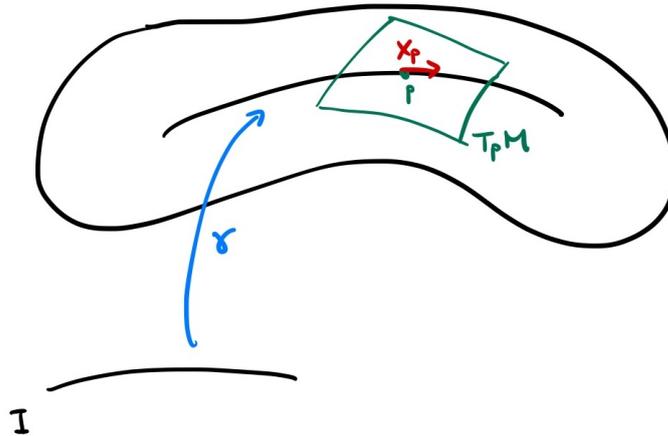


Figure 2.3: Construction of vectors using the directional derivative of a curve at a point.

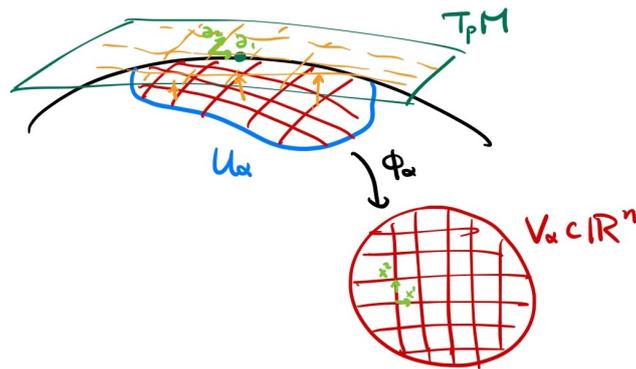


Figure 2.4: Introducing coordinates to the tangent space $T_p M$.

Note that this satisfies two important properties:

- (i) Linearity: $X_p(\alpha f + \beta g) = \alpha X_p(f) + \beta X_p(g)$ with α, β constants.
- (ii) Leibniz rule: $X_p(fg) = X_p(f)g(p) + f(p)X_p(g)$.

However, we have yet to make any statements about X_p furnishing a vector space. At this point, we have only defined individual vectors using the directional derivative of a curve embedded in the manifold. I would like to take a pause here. Let us jump a bit ahead and assumed the vectors do span a vector space. How do we assign coordinates on this vector space? Clearly, a natural basis is given by a chart at p , ϕ_p , and we can project the coordinates of the chart onto the tangent space (see Figure 2.4 for an illustration). To

do this, we introduce the chart $\phi = (x^1, \dots, x^n)$ in the neighbourhood of p . Then we can write,

$$f \circ \gamma = f \circ \phi^{-1} \circ \phi \circ \gamma, \quad (2.3)$$

by inserting an identity map $\text{id} = \phi^{-1} \circ \phi : M \rightarrow M$. Then we have the maps $f \circ \phi^{-1} = \mathbb{R}^n \rightarrow \mathbb{R}$ and $\phi \circ \gamma : I \rightarrow \mathbb{R}^n$ and we write,

$$X_p(f) = \left(\frac{\partial(f \circ \phi^{-1})(x)}{\partial x^\mu} \right)_{\phi(p)} \left(\frac{dx^\mu(\gamma(t))}{dt} \right)_{t=0} = \left(\frac{\partial}{\partial x^\mu} \right)_{\phi(p)} (f) \left(\frac{dx^\mu(\gamma(t))}{dt} \right)_{t=0}. \quad (2.4)$$

For the benefit of our sanity let us drop the action on the function f (assume we are looking at elements in a vector space acting on the space of linear functions), and we have ³,

$$X_p = \left(\frac{dx^\mu(\gamma(t))}{dt} \right)_{t=0} \left(\frac{\partial}{\partial x^\mu} \right)_p. \quad (2.5)$$

This immediately reminds us of the coordinate basis in special relativity — $\frac{\partial}{\partial x^\mu}$ are just the coordinate basis and clearly in this basis we have,

$$X_p^\mu = \left(\frac{dx^\mu(\gamma(t))}{dt} \right)_{t=0}, \quad (2.6)$$

as the components of X_p in this coordinate basis!

We now go back to the question why these tangent vectors span a vector space.

Proposition 2.1. *The set of all tangent vectors at p form an n -dimensional vector space $T_p(M)$ known as the **tangent space at p** .*

Proof. There are two important properties in showing that some algebraic structure is a vector space. Firstly, linearity. To do this we consider two curves γ and λ through p , and we can define addition and multiplication by constant of tangent vectors by defining $\alpha X_p + \beta Y_p$ to be the linear map acting on the space of functions $f \mapsto \alpha X_p(f) + \beta Y_p(f)$. This is indeed a tangent vector to a curve through p , as one can construct a curve using a local chart ϕ as,

$$\nu(t) = \phi^{-1} [\alpha(\phi(\gamma(t)) - \phi(p)) + \beta(\phi(\lambda(t)) - \phi(p)) + \phi(p)], \quad (2.7)$$

and you can check that the tangent vector of $\nu(t)$ is $Z_p = \alpha X_p + \beta Y_p$. The second property is a bit more difficult. We first show there is an n -dimensional basis — this is already explained above, but in the chart ϕ we can pick the curve that only moves in the i -th coordinate, i.e.

$$\lambda_\mu(t) = \phi^{-1}(x^1(p), \dots, x^{\mu-1}(p), x^\mu(p) + t, x^{\mu+1}(p), \dots, x^n(p)), \quad (2.8)$$

so the curve only traverses along a coordinate direction in the chart. This defines the tangent vector to this curve as $\left(\frac{\partial}{\partial x^\mu} \right)_p$. The n tangent vectors are linearly independent as

³I have dropped $\phi(p)$ for p for notational convenience, so we intuitively treat those things as the same thing when we see them together.

the coordinates as linearly independent, and the tangent vectors span the vector space as we can write,

$$X_p = \left(\frac{dx^\mu(\gamma(t))}{dt} \right)_{t=0} \left(\frac{\partial}{\partial x^\mu} \right)_p . \quad (2.5)$$

as we described before. \square

We can now ask what happens when we change the coordinate basis — i.e. switch to a different chart. To do this consider ϕ and ϕ' being different coordinate basis $\{x^\mu\}$ and $\{x'^\mu\}$ respectively. We look at how the basis vector changes. For any smooth function f , we have,

$$\begin{aligned} \left(\frac{\partial}{\partial x^\mu} \right)_p (f) &= \left(\frac{\partial}{\partial x^\mu} (f \circ \phi^{-1}) \right)_{\phi(p)} \\ &= \left(\frac{\partial}{\partial x^\mu} [(f \circ \phi'^{-1}) \circ (\phi' \circ \phi^{-1})] \right)_{\phi(p)} \end{aligned}$$

Note that $f \circ \phi'^{-1}$ is just the function in coordinates x' and $\phi' \circ \phi^{-1}$ is just the functions $x'^\mu(x)$. So using the chain rule, we can write,

$$\begin{aligned} \left(\frac{\partial}{\partial x^\mu} \right)_p (f) &= \left(\frac{\partial}{\partial x^\mu} [(f \circ \phi'^{-1})(x'(x))] \right)_{\phi(p)} \\ &= \left(\frac{\partial x'^\nu}{\partial x^\mu} \right) \left(\frac{\partial}{\partial x'^\nu} \right)_p (f) . \end{aligned}$$

So now we can compare the coordinate basis to get,

$$\left(\frac{\partial}{\partial x^\mu} \right)_p = \left(\frac{\partial x'^\nu}{\partial x^\mu} \right) \left(\frac{\partial}{\partial x'^\nu} \right)_p \quad (2.9)$$

We like transforming the components of the vectors themselves instead. So let us write $X = X^\mu \partial_\mu = X'^\nu \partial'_\nu$, which gives,

$$X = X^\nu \left(\frac{\partial}{\partial x^\nu} \right)_p = X'^\mu \left(\frac{\partial x'^\mu}{\partial x^\nu} \right) \left(\frac{\partial}{\partial x'^\mu} \right)_p , \quad (2.10)$$

and hence we have,

$$X'^\mu = \left(\frac{\partial x'^\mu}{\partial x^\nu} \right)_{\phi(p)} X^\nu . \quad (2.11)$$

This is exactly the transformation rule for the contravariant vectors you have seen in B2! So it is evident that we have found the correct objects — the vectors we have just described are exactly the vectors of contravariant form in the Minkowski sense. Cool!

2.3 Covectors and tensors

We have now found a natural vector space for our manifolds, the tangent space at p , $T_p M$. Now we can use what we have learnt in our linear algebra courses to construct new vector spaces.

1. **Dual spaces.** Recall that for V a real vector space, we can define the **dual space of V** , V^* as the vector space of linear maps from V to \mathbb{R} . To this basis we can define a dual basis f^μ which acts on the basis e_μ of V as $f^\nu(e_\mu) = \delta_{\mu\nu}$. The dual of dual is isomorphic to the original vector space V .

2. **Direct products.** Recall that if V_i , $i = 1, \dots, n$ are vector spaces, the **direct product** is the vector space which the underlying set is the Cartesian product $V_1 \times \dots \times V_n$. In particular, if (v_1^i, \dots, v_n^i) , $(w_1^i, \dots, w_n^i) \in V_1 \times \dots \times V_n$, we will have,

$$(v_1^i, \dots, v_n^i) + (w_1^i, \dots, w_n^i) = (v_1^i + w_1^i, \dots, v_n^i + w_n^i), \quad (2.12)$$

$$\lambda \cdot (v_1^i, \dots, v_n^i) = (\lambda v_1^i, \dots, \lambda v_n^i). \quad (2.13)$$

These elementary facts allow us to construct new objects using $T_p M$. Before we continue however, let us have a look at some multilinear algebra which will be important in our discussion.

Definition 2.9. Let V_1, \dots, V_n and W be vector spaces. A map $f : V_1 \times \dots \times V_n \rightarrow W$ is **multilinear** if it is linear as a function in each variable separately when the others are held fixed. So for each i ,

$$f(v_1, \dots, \alpha v_k + \beta w_k, \dots, v_n) = \alpha f(v_1, \dots, v_k, \dots, v_n) + \beta f(v_1, \dots, w_k, \dots, v_n). \quad (2.14)$$

To construct tensor spaces, we need to think about the vector space of multilinear functions $\text{Maps}(V_1, \dots, V_k; W)$. This space should really be viewed as the set of all linear combination of objects of the form (w_1, \dots, w_n) where $w_1 : V_1 \rightarrow W$ is the dual vector for V_1 . Note that this is different from direct products! The elements of direct products are just the elements of the form (v_1, \dots, v_n) , whereas for tensor products we will now need to define as follows.

Definition 2.10. The **tensor product** $U \otimes V$ is the \mathbb{K} -vector space generated by the symbols $u \otimes v$ with $u \in U$ and $v \in V$, modulo the relations,

$$(\lambda_1 u_1 + \lambda_2 u_2) \otimes v = \lambda_1 (u_1 \otimes v) + \lambda_2 (u_2 \otimes v), \quad (2.15)$$

$$u \otimes (\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 (u \otimes v_1) + \lambda_2 (u \otimes v_2). \quad (2.16)$$

Alternatively, tensor products can be defined as the formal linear combinations of vector spaces, with a general element being in the form,

$$\sum c_i (v_1^i, \dots, v_n^i) \in V_1 \otimes \dots \otimes V_n. \quad (2.17)$$

This abstract way of defining tensor products is perhaps a bit confusing, but in fact there is a canonical isomorphism between the spaces, given by the following proposition,

Proposition 2.2. *If V_1, \dots, V_n are finite-dimensional vector spaces, there is a canonical isomorphism,*

$$V_1^* \otimes \dots \otimes V_n^* \cong \text{Maps}(V_1, \dots, V_n; \mathbb{R}). \quad (2.18)$$

Proof. Key is to use a basis to show the isomorphism, see Proposition 12.10 in [1] for details. \square

Abstract tensor spaces

We should really, really define tensor spaces using formal linear combinations, given by the following definition.

Definition 2.11. Let S be a set. A **formal linear combination of elements of S** is a function $f : S \rightarrow \mathbb{R}$ such that $f(s) = 0$ for all but finitely many $s \in S$. The **free vector space on S** , denoted as $\mathcal{F}(S)$ is the set of all formal linear combination of elements of S . $\mathcal{F}(S)$ is then a vector space over \mathbb{R} under pointwise addition and scalar multiplication.

To construct tensor products, we start with V_1, \dots, V_n real vector spaces and form the free vector space $\mathcal{F}(V_1 \times \dots \times V_n)$ given by the set of all finite formal linear combinations of n -tuples (v_1, \dots, v_n) . Then,

Definition 2.12. The **tensor product space** $V_1 \otimes \dots \otimes V_n$ is the quotient vector space,

$$V_1 \otimes \dots \otimes V_n = \mathcal{F}(V_1 \times \dots \times V_n) / \mathcal{R} \quad (2.19)$$

with the subspace \mathcal{R} generated by elements of the forms,

$$(v_1, \dots, av_k, \dots, v_n) - a(v_1, \dots, v_k, \dots, v_n), \quad (2.20)$$

and

$$(v_1, \dots, v_k + w_k, \dots, v_n) - (v_1, \dots, v_k, \dots, v_n) - (v_1, \dots, w_k, \dots, v_n). \quad (2.21)$$

This is just a formal way of defining tensor spaces, but you will see the mathematical construction $V = A / \sim$ occur very commonly in mathematics and physics — this is just a normal way of constructing a space where we have some additional identifications (here being the ‘normal linearity rules’ of the tensor product).

Enough of this mathematical spill. What can we now do for $T_p M$?

Let us first consider dual spaces. Clearly we can construct some dual space of $T_p M$.

Definition 2.13. The dual space of $T_p M$ is denoted $T_p^* M$ and is called the **cotangent space at p** . An element of this space is called a **covector at p** .

We can similarly construct a coordinate basis for the covector space, but we will first need a notion.

Definition 2.14. Let $f : M \rightarrow \mathbb{R}$ be a smooth function. Define a covector $(df)_p$ called **gradient of f at p** as,

$$(df)_p(X) = X(f) \quad (2.22)$$

for any vector $X \in T_p(M)$.

In a local chart, we can now pick $f = x^\mu$. This means that we can have,

$$(dx^\mu)_p \left(\left(\frac{\partial}{\partial x^\nu} \right)_p \right) = \delta_\nu^\mu, \quad (2.23)$$

and $\{(dx^\mu)_p\}$ is the dual basis of $\left\{ \left(\frac{\partial}{\partial x^\mu} \right)_p \right\}$. You can also check that the gradient of function f now has coordinate basis,

$$[(df)_p]_\mu = \left(\frac{\partial(f \circ \phi^{-1})}{\partial x^\mu} \right)_{\phi(p)}, \quad (2.24)$$

and by changing the coordinate basis you will now get the usual covector transformation rule,

$$(dx^\mu)_p = \left(\frac{\partial x^\mu}{\partial x'^\nu} \right)_{\phi'(p)} (dx'^\nu)_p, \quad (2.25)$$

$$\omega'_\mu = \left(\frac{\partial x^\nu}{\partial x'^\mu} \right)_{\phi'(p)} \omega_\nu. \quad (2.26)$$

Yay! Now let us move on to tensors. Following our discussion of tensor spaces, we can now define tensors as a multilinear map as follows.

Definition 2.15. A tensor of type (p, q) at p is the multilinear map,

$$T : T_p^*M \times \cdots \times T_p^*M \times T_pM \times \cdots \times T_pM \rightarrow \mathbb{R}, \quad (2.27)$$

with p factors of T_p^*M and q factors of T_pM .

Note that the indices are swapped — for every upper (contravariant) index we want a map $T_p^*M \rightarrow \mathbb{R}$ as the dual space of the dual is isomorphic to itself. Similarly for every lower (covariant) index we want a map $T_pM \rightarrow \mathbb{R}$. By introducing a basis $\{e_\mu\}$ for T_pM and dual basis $\{f^\mu\}$ for T_p^*M we will have the components of T as,

$$T^{\mu_1 \cdots \mu_p}_{\nu_1 \cdots \nu_q} = T(f^{\mu_1}, \dots, f^{\mu_p}, e_{\nu_1}, \dots, e_{\nu_q}). \quad (2.28)$$

Let me introduce briefly here what **abstract index notation** is. So far we have used Greek letters μ, ν, \dots to denote the components of a tensor (vectors and covectors are just $(1, 0)$ - and $(0, 1)$ -tensors respectively). The abstract index notation replaces the Greek letters with Latin letters — it simply tells you the object is a tensor of a certain type. This avoids the fact that sometimes we might have $X^1 = 0$ in some basis but $X'^1 \neq 0$ in a different basis. So from now on when you see T^a you should just treat this as, ‘ah, this is a vector!’

2.4 Tensor fields

Well, we have really only defined tensors at a point haven’t we. But in physics we really want to understand how these things vary in spacetime — so we will need to vary them across the manifold. So we will need some notion for that. Let’s start with a vector field.

Definition 2.16. A **vector field** is a map X which maps any point $p \in M$ to a vector $X_p \in T_p M$. This allows us to define a new function $X(f) : M \rightarrow \mathbb{R}$ with $X(f) : p \mapsto X_p(f)$. The vector field is **smooth** if $X(f)$ is a smooth function for any smooth f .

This can be easily generalised to covectors and tensors.

Definition 2.17. A **covector field** is a map ω which maps any point $p \in M$ to a covector $\omega_p \in T_p^* M$. Similar to the vector case, with a vector field and covector field we can define the function $\omega(X) : M \rightarrow \mathbb{R}$ by $\omega(X) : p \mapsto \omega_p(X_p)$. The covector field is **smooth** if $\omega(X)$ is a smooth function for any smooth X .

The principle is thus — to translate the idea of smoothness we start with some smooth function f (which we can use local coordinates to analyse the properties), and then define new functions by composing $X(f)$ for vectors and then $\omega(X)$ for covectors. Neat. This then gives the definition of tensor fields.

Definition 2.18. A **tensor field** is a map T which maps any point $p \in M$ to a tensor T_p at p . Again we define a function $T(\eta_1, \dots, \eta_p, X_1, \dots, X_q) : M \rightarrow \mathbb{R}$. The tensor field is **smooth** if $T(X)$ is a smooth function for any collections of smooth η_i and X^i .

Bundles

The neat notion for defining tensor fields is actually to use the idea of bundles. Let me give you a short introduction here, and only illustrate for vector bundles.

Definition 2.19. A **diffeomorphism** between two manifolds $f : M \rightarrow N$ is a smooth map with a smooth two-sided inverse.

Definition 2.20. A **vector bundle of rank k over manifold X** , $E \rightarrow X$ is a manifold equipped with the following data.

- (i) Call E is the **total space**, X is the **base space**.
- (ii) A smooth surjection map $\pi : E \rightarrow X$ called a **projection**.
- (iii) An open cover $\{U_\alpha\}$ of X such that for each α there is a diffeomorphism called **local trivialisations**,

$$\Phi_\alpha : \pi^{-1}(U_\alpha) \rightarrow U_\alpha \times \mathbb{R}^k, \quad (2.29)$$

such that $\text{pr}_1 \circ \Phi_\alpha = \pi$ (pr_1 means projecting to the first item in the map), and that for all α and β the map $\Phi_\beta \circ \Phi_\alpha^{-1}$ has the form,

$$(U_\alpha \cap U_\beta) \times \mathbb{R}^k \rightarrow (U_\alpha \cap U_\beta) \times \mathbb{R}^k \quad (2.30)$$

with $x \in \mathbb{R}^k$,

$$(b, x) \mapsto (b, g_{\beta\alpha}(b)(x)), \quad (2.31)$$

for some smooth map $g_{\alpha\beta} : U_\alpha \cap U_\beta \rightarrow GL(k, \mathbb{R})$ called **transition functions**.

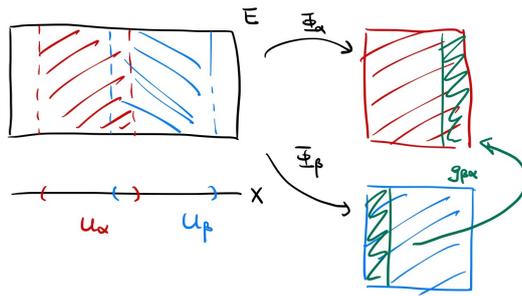


Figure 2.5: An illustration for a vector bundle.

Effectively, a vector bundle is just a way of organising vector data on a manifold (see Figure 2.5). The transition functions here are very similar to the transition functions we have seen for patches (or charts) — they basically just tells us how to go from one description of vector space to a different description of a vector space. In particular, this will allow some twisting of the vector space — the vector space is no longer necessarily just a product space on top of the manifold $E = X \times \mathbb{R}^k$, but we can twist the \mathbb{R}^k in some weird way that admits this smooth patching using the transition functions.

With this we can define the vector fields as follows.

Definition 2.21. A **local section** of a vector bundle $E \rightarrow X$ is the map $s_\alpha : U_\alpha \rightarrow E$ such that $\pi \circ s_\alpha = \text{id}_X$. A **global section** of a vector bundle $E \rightarrow X$ is the map $s : X \rightarrow E$ such that $\pi \circ s = \text{id}_X$.

Definition 2.22. A **vector field on X** is a global section on the vector bundle $E \rightarrow X$.

Crazy how we can describe objects with geometrical notions! We will revisit this when we talk about connections.

3 Metrics

We have discussed how important metrics are in describing gravity in §1. So we should really define what they are, as a mathematical object.

3.1 A first look at metrics

Fundamentally, a metric is just a specific type of tensor field.

Definition 3.1. A **metric tensor** at $p \in M$ is a $(0,2)$ -tensor g with the following properties:

1. Symmetry: $g(X, Y) = g(Y, X)$ for all $X, Y \in T_p(M)$.

2. Non-degeneracy $g(X, Y) = 0$ for all $Y \in T_p(M)$ iff $X = 0$.

I like to think of the reason why a metric tensor needs to vectors as follows. Imagine you are at some point on the manifold and wants to measure the distance between two points far away. You will, at first instance, need to find a way to go to those two points (since things can only be measured locally, as discussed) so we will need to find the curves that will allow us to traverse to those two points. Hence we will get two vectors! Of course, it does not matter which point we go first (hence the symmetry requirement); nor should this intrinsic notion of distance be defined when I cannot reach there via a path (i.e. when $X = 0$, which justifies the non-degeneracy requirement).

The metric is a local construction. In particular, the construction only depends on local data — the vectors are taken from local tangent space. In other words, the metric is an inner product of the tangent space at a point. This is in contrast to the notion of metrics in a metric space.

Metric space

A metric space is a space X together with something known as the distance function, or metric, defined as follows.

Definition 3.2. A **metric space** is the pair (X, d) where X is a non-empty set and $d : X \times X \rightarrow \mathbb{R}$ is the **metric** that satisfies the following properties.

- (i) Non-degeneracy: for $x, y \in X$, $d(x, y) \geq 0$ and $d(x, y) = 0$ iff $x = y$.
- (ii) Symmetry: for $x, y \in X$, $d(y, x) = d(x, y)$.
- (iii) Triangle inequality: for $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$.

Notice that the triangle inequality is not included in the definition in the metric tensor. For manifolds, the construction of metric is local, whereas the distance function in metric spaces $d : X \times X \rightarrow \mathbb{R}$ is global. In other words, the triangle inequality requires comparing the distances between of three points in the space, which may or may not be possible to be described within the same local patch in a manifold.

3.2 Signature of the metric

We want to now look at an important property about the metric. In particular, we can treat g , locally in some coordinates, as a quadratic form which admits a matrix representation. We know from elementary linear algebra that we can diagonalise the matrix using some non-singular matrix S to give $A = S^T D S$ where D is a diagonal matrix with only $\pm 1, 0$ along the diagonal. Then we have the following principle.

Theorem 3.1 (Sylvester's Law of Inertia). *The number of ± 1 and 0 in the diagonal matrix D for a quadratic form A is constant.*

Proof. Use the decomposition to separate the vector space into subspaces where A is positive-definite, negative-definite and null,

$$V = V_+ \oplus V_- \oplus V_0 \tag{3.1}$$

Then since the equality $\dim(V) = \dim(V_+) + \dim(V_-) + \dim(V_0)$, is preserved when we change the basis, the number of positive, negative and zero eigenvalues stay the same. \square

The metric g is a non-degenerate quadratic form. This means that g is only specified by two integers (p, q) , with p and q indicating the number of positive and negative eigenvalues in the matrix representation.

Definition 3.3. The **signature** of a metric is the pair of integers (p, q) with $p + q = n$, indicating the number of $+1$ and -1 in the diagonal matrix.

With a metric, we can now define new kinds of manifolds.

Definition 3.4. A **Riemannian manifold** is an n -manifold M equipped with a metric g such that g has signature $(n, 0)$. A **psuedo-Riemannian manifold** is an n -manifold equipped with an everywhere non-degenerate metric g with signature (p, q) , $p + q = n$.

A psuedo-Riemannian manifold is basically a more general class of manifolds that include Riemannian manifolds. In general relativity we are in particular interested in **Lorentzian manifolds** which are pseudo-Riemannian manifolds with metrics of signature $(n, 1)$.

3.3 Embedding metrics

We have looked at embeddings in the first example sheet. Let me define a bit more clear what that means.

Under construction

I will come back to this.

4 Connections and Geodesics

Previously we have defined vectors using directional derivatives of curves in the manifold. In particular, this requires us to localise at a point on the manifold $p \in M$ so we have defined vectors as members of some tangent space T_pM . Vector fields are just a collection of vectors at different points of the manifold — we single out one member of $v \in T_pM$ for all points $p \in M$.

But we do not want to stop there. In particular, to define differentiation we will need to know how to look at differences between vectors at nearby points, say p and $p' = p + \delta p$. However, we know they live in different vector spaces T_pM and $T_{p'}M$. How do we take differences of objects that are formulated in different spaces?

This is where connection comes in — connections are additional mathematical data needed to fully describe differentiation along curves in manifolds. I will try and develop this in different levels of abstraction, as well as discuss how geodesics, the curves that represent the shortest path between two points.

4.1 Connections I — as a way to fix differences of vectors between two points

Let us make the discussion above explicit and highlight the problem of differences of vectors in curved spacetime. Suppose A^a and $A^a + dA^a$ be the vector at p and p' very close by to p in M . Clearly, we want to look at the quantity dA^a . If this is a vector, then we are done — the problem we have mentioned does not exist! To check this, we introduce some local coordinates in the patch that covers both points x^μ and consider some coordinate transformation to x'^μ . Clearly, the components of A^α transforms as,

$$A'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} A^\nu, \quad (4.1)$$

so a straightforward computation yields,

$$d(A'^\mu) = \frac{\partial x'^\mu}{\partial x^\nu} dA^\nu + \frac{\partial^2 x'^\mu}{\partial x^\nu \partial x^\rho} dx^\rho A^\nu. \quad (4.2)$$

The second term shows that dA^a does not transform like a vector — it is in general non-zero unless the geometry is flat ⁴. So we are screwed. This operation clearly does not work.

Let us propose a way to fix this. In anticipation of what to come, let us suppose we want to define some notion of difference of vectors ∇A^a which transforms like a vector. Clearly we will need to add some corrections,

$$\nabla A^a = dA^a + \delta A^a, \quad (4.3)$$

where δA^a is chosen such that the not very nice bit that makes dA^a not a vector in Eq. (4.2) is cancelled out. What can we add?

1. δA^a must depend on coordinates. Otherwise, the coordinate-dependent correction will not be cancelled out.
2. δA^a must depend on itself since the correction is A^a dependent. Crucially this dependence must be linear. If we have two vectors A^a and B^a , this correction should not change by exchanging the orders of adding the two vectors up $A^a + B^a$ and changing coordinates (and hence effecting the change).

This suggests that we should write down something of the form,

$$\delta A^a = \Gamma_{bc}^a A^b dx^c. \quad (4.4)$$

We have added in Γ_{bc}^a by simply matching indices. This is our first definition of what connection is.

Definition 4.1. The **affine connection (Christoffel symbols)** is a set of n^3 numbers that enforces the vector transformation property of Δv^a .

⁴In this case we will be able to take derivatives! This is the reason why in special relativity (B2) you were able to derivatives of vectors using partial derivatives — mathematically, the tangent spaces of the vectors can be trivially connected by a linear transformation. In other words, the tangent bundle is trivial.

Crucially we have,

$$\nabla A^a = dA^a + \Gamma_{bc}^a A^b dx^c . \quad (4.5)$$

This eerily looks a bit like the covariant derivative. Nice!

What about covectors? Can we also look at the transformation rules for that? There is a trick to this — firstly, note that for scalars the transformation rule is preserved and $\delta f = 0$ for any scalar field f . How do we construct a scalar out of vectors? Clearly we can contract that with a covector,

$$\delta(A^a B_a) = A^a \delta B_a + B_a \delta A^a = 0 . \quad (4.6)$$

Now we can substitute the transformation rule we have before to obtain,

$$0 = A^a (\delta B_a + \Gamma_{ab}^c dx^b B_c) . \quad (4.7)$$

Since A^a is arbitrary, we then have $\delta B_a = -\Gamma_{ab}^c dx^b B_c$ we obtain,

$$\nabla B_a = dB_a - \Gamma_{ab}^c dx^b B_c . \quad (4.8)$$

We then obtain the covariant derivative rule for covectors.

4.2 Connections II — as transformations of tetrads

The argument in the above subsection is a bit too intuitive. We have hand-added in some numbers to solve a problem without real mathematical justification as to why it works. Could we perhaps understand connections in a different way that is a bit more mathematically-justified?

To do this let me first introduce a notion called tetrads. Remember that in a manifold we can locally take some coordinates in a local patch which gives a coordinate basis ∂^μ . This basis is however generically not orthonormal, but in tensor calculations orthonormal basis are very useful. So let us consider some non-coordinate, orthonormal basis of vector fields $\{e_\mu^a\}$. This gives the following definition.

Definition 4.2. A **tetrad** is an orthonormal basis of vector fields, constructed locally in a patch that satisfies,

$$g_{ab}(e_\mu)^a(e_\nu)^b = (e_\mu)^a(e_\nu)_a = \eta_{\mu\nu} , \quad (4.9)$$

where $\eta_{\mu\nu}$ is the Minkowski metric. This also satisfies,

$$\eta^{\mu\nu}(e_\mu)^b(e_\nu)_a = \delta_a^b , \quad (4.10)$$

with δ_a^b being the identity map on vectors.

We can think of these vectors as basis vectors. In particular, the following properties follow from our definition.

Lemma 4.1. *The orthonormal basis of vector field satisfies,*

$$\eta_{\mu\nu} e_a^\mu e_b^\nu = g_{ab} , \quad \eta^{\mu\nu} (e_\mu)^b (e_\nu)_a = \delta_a^b , \quad (4.11)$$

Proof. To see this, contract the first equation with e_ρ^b on both sides,

$$\eta_{\mu\nu} e_a^\mu e_b^\nu e_\rho^b = \eta_{\mu\nu} e_a^\mu \delta_\rho^\nu = (e_\rho)_a = g_{ab} e_\rho^b. \quad (4.12)$$

e_ρ^b is arbitrary so the first equation holds true. We can also write this equation as,

$$e_a^\mu (e_\mu)_b = g_{ab}, \quad (4.13)$$

so raising the b index gives the second equation. \square

Another way to think about tetrads is to think of them as coordinate basis vectors, i.e. the covectors can be defined as,

$$(e_\mu)^a = \lim_{\delta x^\mu \rightarrow 0} \frac{\delta s^a}{\delta x^\mu} \quad (4.14)$$

where δs is the vector that encodes the infinitesimal interval. In a local coordinate patch, these vectors are exactly the directional derivatives of the curve defined by the coordinate line x^μ at p . The a -index here is an abstract tensor index — it indeed transforms like a tensor under a change of coordinates, but the extra μ is not a tensor index. The coordinate index μ here is related to how we have constructed the vectors, and we have defined that to be related to the Minkowski metric in a local patch. This allows us to raise and lower the index using the metric. Sometimes I will also not write the μ symbol and implicitly assume that one can find this by going to a local coordinate patch to specify the coordinate lines of x^μ .

We can now think about how to move coordinate basis vectors to the next point. In particular I can write,

$$e_a(p') = e_a(p) + \delta e_a. \quad (4.15)$$

Clearly, $\frac{\partial e_a}{\partial x^\mu}$ may not lie in the tangent vector space $T_{p'}M$. We can project this vector to this tangent space, and expand using the local basis at p' ,

$$\frac{\partial e_a}{\partial x^\mu} = \Gamma_{a\mu}^b e_b. \quad (4.16)$$

In other words, we have defined some coefficient Γ_{ac}^b that connects the basis at two nearby points p and p' . This motivates the definition below ⁵.

Definition 4.3. The **affine connections (Christoffel symbols)** are the n^3 -coefficients Γ_{ac}^b defined as the change of the basis vector,

$$\Gamma_{ac}^b = e_\mu^b \partial_c e_a^\mu. \quad (4.17)$$

⁵Note that in the definition I have replaced the coordinate index μ with an abstract index c . I want to emphasise that this is for notational convenience, and the two lower indices in the Christoffel symbols are indeed of different nature. We will see why this is the case when we discuss covariant derivatives, but we can note now that one of the indices is intrinsically related to a vector index, where is the other one is related to a coordinate.

Note that by noting that the reciprocity relation $e_\mu^a e_b^\mu = \delta_b^a$, we will have,

$$\partial_c(e_\mu^a e_b^\mu) = 0 = \partial_c(e_\mu^a) e_b^\mu + e_\mu^a \partial_c(e_b^\mu) \quad (4.18)$$

so therefore we can also define the affine connections using,

$$\partial_c e^a = -\Gamma_{bc}^a e^b. \quad (4.19)$$

It is now obvious to derive the transformation properties of the affine connection. In particular, we can use the fact that in the primed coordinates,

$$\Gamma_{bc}^{\prime a} = e_\mu^{\prime a} \frac{\partial e_b^{\prime \mu}}{\partial x^{\prime c}} \quad (4.20)$$

and using the fact that,

$$e^{\prime a} = \frac{\partial x^{\prime a}}{\partial x^b} e^b. \quad (4.21)$$

you should be able to derive,

$$\Gamma_{bc}^{\prime a} = \frac{\partial x^{\prime a}}{\partial x^d} \frac{\partial x^f}{\partial x^{\prime b}} \frac{\partial x^g}{\partial x^{\prime c}} \Gamma_{fg}^d + \frac{\partial x^{\prime a}}{\partial x^d} \frac{\partial^2 x^d}{\partial x^{\prime c} \partial x^{\prime b}}. \quad (4.22)$$

Transformation Rules

A cheat way to look at tensor transformation rules under a coordinate change $x \mapsto x'$ is to match indices. In particular, think of the coordinates as always having an upstairs index, so we always need to transform that index by contracting that with the correct ∂x^a either in the denominator or numerator. Therefore,

$$V^{\prime a} = \frac{\partial x^{\prime a}}{\partial x^b} V^b, \quad (4.23)$$

$$\omega'_a = \frac{\partial x^b}{\partial x^{\prime a}} \omega^b. \quad (4.24)$$

and so on. This gives a quick way of writing transformation rules down.

In particular, we can now define covariant derivatives as before. We look at the total derivative,

$$\frac{d}{dx^a}(v^b e_b) = (\partial_a v^b) e_b + v^b \partial_a e_b = \left(\partial_a v^b + \Gamma_{ca}^b v^c \right) e_b \quad (4.25)$$

in which we define the last term as the covariant derivative, explicitly,

$$\nabla_a v^b = \partial_a v^b + \Gamma_{ca}^b v^c. \quad (4.26)$$

Such a covariant derivative I will, for the time being, denote as ‘the associative covariant derivative to the affine connection Γ_{bc}^a . One can of course do that for the $(0,1)$ -tensors (1-forms) as well, so let me leave that as an exercise.

So far we haven’t said anything about how the connections and the metric are related. In particular, there is a continuum of different connections we can actually choose which satisfies the property in Eq. (4.17). It will become clear why we have chosen this, but we will isolate a specific set of connections as per the following definition.

Definition 4.4. The **Levi-Civita connection** is the metric-compatible connection satisfying the following properties.

- (i) The connection is torsionless, i.e. the **torsion tensor** defined as,

$$T_{ac}^b = \Gamma_{ac}^b - \Gamma_{ca}^b, \quad (4.27)$$

is identically zero.

- (ii) The Christoffel symbols are given by the metric as,

$$\Gamma_{bc}^a = \frac{1}{2}g^{ad}(\partial_b g_{dc} + \partial_c g_{bc} - \partial_d g_{bc}). \quad (4.28)$$

This can be derived from differentiating the tetrad relation, $g_{ab} = e_a^\mu(e_\mu)_b$ ⁶, and the condition is equivalent to the associated covariant derivative to satisfy,

$$\nabla_a g_{bc} = 0. \quad (4.29)$$

Such a choice of connection always exists and is unique for pseudo-Riemannian manifolds.

We will go back to the last point after discussing what covariant derivatives are. But from now on when we write down connection coefficients you can always assume that we are using the metric-compatible connection coefficients such that the bottom indices are symmetric and the covariant derivative of the metric is zero.

4.3 Intrinsic derivatives and geodesics

Before we begin, let us go back to curves. We have previously defined curves as maps from some interval into the manifold M . An important class of curves is generated by vector fields — these are known as **integral curves**.

Definition 4.5. Let $v \in \mathfrak{X}(M)$ be a vector field on M ⁷. An **integral curve of v in M through p** is the curve through $p \in M$ such that the tangent vector at every point is given by v .

For some parameter parametrising the curve λ , the integral curve is given by, in some local coordinate chart,

$$\frac{dx^\mu}{d\lambda} = X^\mu(x(\lambda)), \quad x^\mu(0) = x_p^\mu. \quad (4.30)$$

Now that an integral curve is defined, we can think about some derivative along the curve γ . This is known as the intrinsic derivative given by the following definition.

Definition 4.6. Let γ be the integral curve on the manifold generated by some vector field $v \in \mathfrak{X}(M)$. The **intrinsic derivative of the integral curve γ** is defined as the derivative along the vector field,

$$\frac{D}{D\lambda} = v^a \nabla_a. \quad (4.31)$$

⁶This is effectively what you have done in Q7 of the first sheet!

⁷Please don't ask me why the set of vector fields on M is denoted as $\mathfrak{X}(M)$. This is some natural convention normally taken by mathematicians.

The intrinsic derivative can be understood as follows. Suppose we take the derivative of some vector field w ,

$$\frac{d}{d\lambda}(w) = \frac{d}{d\lambda}(w^a e_a), \quad (4.32)$$

this is indeed very similar to what we were doing before! The upshot is we can use the chain rule and write,

$$\frac{dx^a}{d\lambda} \frac{d}{dx^a}(w^b e_b) = v^a (\nabla_a w^b) e_b, \quad (4.33)$$

since the tangent vector is just the vector field v . This establishes the relationship in Eq. (4.31).

We can now discuss geodesics! There are two equivalent definitions of geodesics which are related. We start with the first one, intrinsically related to tangent vectors of the curve.

Definition 4.7. An **affine geodesic** is a curve $\gamma : I \rightarrow M$ whose tangent vector is constant along the curve.

The definition can be motivated from looking at geodesics in Euclidean space. It is clear that the geodesics, i.e. straight lines, have their tangent vectors always pointing along the line in Euclidean space. Note however that this does not stop the magnitude of the tangent vectors changing along the path! This means that along the path parametrised by some parameter λ , the tangent vector \mathbf{u} must be effected by a change at most proportional to itself, i.e.

$$d\mathbf{u} = f(\lambda)\mathbf{u} d\lambda. \quad (4.34)$$

In terms of intrinsic derivatives, we can write the change of the tangent vector component as,

$$\frac{Du^a}{D\lambda} = f(\lambda)u^a. \quad (4.35)$$

This is the geodesic equation! Recalling that the intrinsic derivative of an integral curve can be swapped with the derivative $u^a \nabla_a$, we can write the following definition.

Definition 4.8. The **geodesic equation** is given by,

$$u^a \nabla_a u^b = f(\lambda)u^b. \quad (4.36)$$

One can choose a parametrisation such that $f(\lambda) = 0$, in which case the geodesic is known as **affinely-parametrised**, and satisfies the **affinely-parametrised geodesic equation**,

$$u^a \nabla_a u^b = 0. \quad (4.37)$$

Before we continue, let us look at the coordinate form of the affinely-parametrised geodesic equation. Using the fact that,

$$u^a = \frac{dx^a}{d\lambda}, \quad (4.38)$$

and the expression for the covariant derivative, we get,

$$\frac{dx^a}{d\lambda} \nabla_a \left(\frac{dx^b}{d\lambda} \right) = \frac{dx^a}{d\lambda} \left(\frac{d\lambda}{dx^a} \right) \frac{d^2 x^b}{d\lambda^2} + \Gamma_{ca}^b \left(\frac{dx^c}{d\lambda} \right) \left(\frac{dx^a}{d\lambda} \right) = 0. \quad (4.39)$$

Simplifying the equation and relabelling indices gives,

$$\frac{d^2 x^a}{d\lambda^2} + \Gamma_{bc}^a \frac{dx^b}{d\lambda} \frac{dx^c}{d\lambda} = 0, \quad (4.40)$$

which is just the usual geodesic equation in coordinate form!

Let us return to the question about parametrisations. An affine parametrisation (of an affine geodesic) is an advantaged parametrisation which is possible under some circumstances in pseudo-Riemannian manifolds. To see this, we can pick the affinely-parametrised curve and reparametrise using $\lambda \mapsto \lambda'(\lambda)$. The coordinate geodesic equation then becomes,

$$\frac{d^2 x^a}{d\lambda'^2} + \Gamma_{bc}^a \frac{dx^b}{d\lambda'} \frac{dx^c}{d\lambda'} = \left(\frac{d^2 \lambda}{d\lambda'^2} \right) \left(\frac{d\lambda'}{d\lambda} \right) \frac{dx^a}{d\lambda}. \quad (4.41)$$

This has the following consequences. Firstly, by comparing with the non-affinely parametrised geodesic equation in coordinate form, we can identify,

$$f(\lambda') = \left(\frac{d^2 \lambda}{d\lambda'^2} \right) \left(\frac{d\lambda'}{d\lambda} \right). \quad (4.42)$$

Therefore, if we can find λ' such that this is integrable — and clearly it is and we can find the function,

$$\lambda(\lambda') = \int^{\lambda'} d\mu e^{\int^{\mu} dx f(x)}. \quad (4.43)$$

Therefore, given $f(\lambda')$, we can always find a reparametrisation $\lambda(\lambda')$ to make it into an affinely-parametrised geodesic.

This parametrisation is affine as any reparametrisation $\lambda \mapsto a\lambda + b$ where a and b are constants will leave the affinely-parametrised geodesic equation the same (owing to the double derivative). In fact, the following definition might be useful to motivate the word affine...

Definition 4.9. An **affine transformation** is a transformation that preserves collinearity (all points lying on a line still lie on the same line) and ratios of the distances.

Affine transformations are roughly linear transformations with translations. The name for affine connections comes from identifying tangent spaces of Euclidean space by translation — hence making Euclidean space into an affine space (roughly, a space without an origin). I won't go too deep into this but if you are interested you can read about this in [2].

4.4 Variational methods and geodesics

There is a second definition of a geodesic.

Definition 4.10. A **metric geodesic** is a curve $\gamma : I \rightarrow M$ whose length is stationary with respect to small variations in the path.

This is where the normal notion of geodesic — the path of shortest distance between two points come from. One can perhaps ask why there are two notions of geodesics. The two notions are intrinsically different, and on a manifold where a torsion-free connection exists (i.e. the torsion tensor defined with respect to an affine connection in Eq.(4.27) is zero), we have the following theorem.

Proposition 4.1. *On a manifold where a torsion-free connection exists, the two notions of geodesics are the same, i.e. the curve that extremises the length is the curve that leaves the tangent vector in the same direction.*

Proof. See how the derivation of the geodesic equation arises from extremising the action below. We will need to use the existence of the Levi-Civita connection. \square

Let us proceed to see how it works.

Variational methods

Here is a quick reminder of how actions and variational methods work. We start with a functional called an **action** which is the integral of some **Lagrangian**,

$$S[x(t)] = \int_{t_i}^{t_f} dt L(x_i(t), \dot{x}_i(t)) . \quad (4.44)$$

To extremise the action, we find a point where $\delta S = 0$ with the end points held fixed, $\delta x_i(t_i) = \delta x_i(t_f) = 0$. Here we consider, for the simple case where the Lagrangian only depends on the functions $x_i(t)$ and its derivatives but has no explicit dependence on the parameter t . Varying the action then gives,

$$\delta S = \int_{t_i}^{t_f} dt \left[\frac{\partial L}{\partial x_i} \delta x_i + \frac{\partial L}{\partial \dot{x}_i} \delta \dot{x}_i \right] , \quad (4.45)$$

using integration by parts we have,

$$\delta S = \int_{t_i}^{t_f} dt \delta x \left[\frac{\partial L}{\partial x_i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_i} \right) \right] + \left[\frac{\partial L}{\partial \dot{x}_i} \delta x_i \right]_{t_i}^{t_f} . \quad (4.46)$$

Clearly the last term vanishes because of the boundary conditions $\delta x_i(t_i) = \delta x_i(t_f) = 0$. Since the variation $\delta x_i(t)$ is arbitrary, we have the **Euler-Lagrange equations**,

$$\frac{\partial L}{\partial x_i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_i} \right) = 0 , \quad \forall x_i . \quad (4.47)$$

We need to set-up what kind of Lagrangian we would like to use — in particular, we need to pick a quantity that an observer travelling in curved spacetime will be able to measure. The clear quantity here is the **proper time**, defined as,

$$ds^2 = -d\tau^2 . \quad (4.48)$$

Therefore, we write our action as the total proper time,

$$S = \int d\tau = \int_0^1 d\lambda \sqrt{-g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu}, \quad (4.49)$$

where $x = x(\lambda)$ and λ is a parameter on the curve (with 0 and 1 being the end points of the curve) and the dot corresponds to taking a derivative with respect to λ . Let us label

$$L = \sqrt{-g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu}, \quad (4.50)$$

for notational convenience, which means the Euler-Lagrange equations give,

$$\frac{d}{d\lambda} \left(\frac{\partial L}{\partial \dot{x}^\mu} \right) - \frac{\partial L}{\partial x^\mu} = 0. \quad (4.51)$$

A straightforward computation gives,

$$\frac{\partial L}{\partial \dot{x}^\mu} = -\frac{1}{L} g_{\mu\nu} \dot{x}^\nu, \quad (4.52)$$

$$\frac{\partial L}{\partial x^\mu} = -\frac{1}{2L} \partial_\mu g_{\nu\rho} \dot{x}^\nu \dot{x}^\rho, \quad (4.53)$$

For convenience, we can change the differentials by using the fact that $(\frac{d\tau}{d\lambda})^2 = L^2$, which gives ⁸,

$$\frac{d}{d\tau} \left(g_{\mu\nu} \frac{dx^\nu}{d\tau} \right) - \frac{1}{2} g_{\nu\rho,\mu} \dot{x}^\nu \dot{x}^\rho = 0. \quad (4.54)$$

Here I have used,

$$g_{\nu\rho,\mu} := \partial_\mu g_{\nu\rho}. \quad (4.55)$$

Now we can proceed and take the differential to give,

$$g_{\mu\nu} \frac{d^2 x^\nu}{d\tau^2} + g_{\mu\nu,\rho} \frac{dx^\rho}{d\tau} \frac{dx^\nu}{d\tau} - \frac{1}{2} g_{\nu\rho,\mu} \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau} = 0, \quad (4.56)$$

The second term is symmetrised with respect to ρ and ν and looks very similar to the first two terms in the Levi-Civita connection! We then just multiply by the inverse metric to get,

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma_{\nu\rho}^\mu \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau} = 0, \quad (4.57)$$

with $\Gamma_{\nu\rho}^\mu$ the Levi-Civita connection coefficients, given by,

$$\Gamma_{\nu\rho}^\mu = \frac{1}{2} g^{\mu\sigma} (g_{\sigma\nu,\rho} + g_{\sigma\rho,\nu} - g_{\nu\rho,\sigma}). \quad (4.58)$$

This is where the metric connection comes in. We have obtained the same geodesic equation as before (we have already addressed the parametrisation issue above). In particular, notice that we used the Levi-Civita connection which is torsionless. If the Levi-Civita does not

⁸By using this parametrisation, we have actually picked an affine parameter. For the general case, you can check that there is an extra term on the RHS of the equation which gives the non-affinely parametrised geodesic equation, see §3.18 of [3].

exist, or that we have used a different set of connection coefficients in our derivation that is not torsion-free, we will instead of the relation ⁹

$$\frac{1}{2}(g_{\sigma\nu,\rho} + g_{\sigma\rho,\nu} - g_{\nu\rho,\sigma}) - \Gamma_{(\nu\rho)}^{\mu}g_{\mu\sigma} + \Gamma_{[\nu\sigma]}^{\mu}g_{\mu\rho} + \Gamma_{[\rho\sigma]}^{\mu}g_{\mu\nu} = 0, \quad (4.59)$$

so our calculation will not reduce to the simple form as we have stated.

This is nice! We have reduced our problem to computing variations of a Lagrangian. How neat that everything lines up in the end!

EL equations for Γ_{bc}^a

In fact, a super important technique in obtaining the Levi-civita connection components is to use the Euler-Lagrange equations. Since squaring and rescaling the Lagrangian does not change the geodesic equations as the Euler-Lagrange equations (up to parametrisation), we can use the Lagrangian,

$$L = -\frac{1}{2}g_{\mu\nu}(x(\tau))\frac{dx^{\mu}}{d\tau}\frac{dx^{\nu}}{d\tau} \quad (4.60)$$

which allows one to directly compute the connection coefficients. Let us look at an example.

Example 4.1. Suppose we look at the flat 2d spherical metric,

$$ds^2 = dr^2 + r^2d\theta^2. \quad (4.61)$$

In this case, we write the Lagrangian as,

$$L = \frac{1}{2}(\dot{r}^2 + r^2\dot{\theta}^2), \quad (4.62)$$

where the dot indicates the derivative with respect to proper time. The Euler-Lagrange equations give the geodesic equations, which are,

$$r : \ddot{r} - r\dot{\theta}^2 = 0, \quad (4.63)$$

$$\theta : \ddot{\theta} + \frac{2}{r}\dot{r}\dot{\theta} = 0. \quad (4.64)$$

From this we can directly read out the non-zero connection coefficients — $\Gamma_{\theta\theta}^r = -r$ and $\Gamma_{r\theta}^{\theta} = \frac{1}{r}$. The rules are this — upper index of Γ_{bc}^a indicates the coordinate of the EL equation, the bottom indices indicate what variables with single dots \dot{r} , $\dot{\theta}$ the coefficient is related to, and if $b \neq c$ we divide by a factor of 2 as there are two contributions in the geodesic equation.

⁹Recall you have done this in Q7 of the first sheet.

4.5 Covariant derivatives and Koszul connections

5 Curvature

A Maps between manifolds

Once we have manifolds, it is natural to consider the maps between them. Let us consider two manifolds, M and N , and a map between them $f : M \rightarrow N$.

Definition A.1. The map $f : M \rightarrow N$ is **smooth** iff for all charts φ_α and φ_a of M and N respectively, $\varphi_a \circ f \circ \varphi_\alpha^{-1}$ as maps between $\mathbb{R}^{\dim n_M}$ and $\mathbb{R}^{\dim n_N}$.

In topology we call maps that have a continuous inverse homeomorphism. Clearly we should have an analogue here.

Definition A.2. A **diffeomorphism** $f : M \rightarrow N$ is a smooth map with a two-sided inverse. Two manifolds are diffeomorphic if there exists a diffeomorphism between them.

References

- [1] J. Lee, *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics, Springer, 2003.
- [2] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry, Volume 1*. A Wiley Publication in Applied Statistics, Wiley, 1996.
- [3] M. P. Hobson, G. P. Efstathiou, and A. N. Lasenby, *General relativity: An introduction for physicists*. 2006.